Adversarial event generator tuning with Bayesian Optimization

Maxim Borisyak, Andrey Ustyuzhanin

National Research University Higher School of Economics (HSE)

July 7, 2018

Event Generator Tuning

Intro

We consider problem of tuning parameters of event generators to 'real' data:

- generating samples is expensive;
- generator is non-differentiable.

Working example: `Pythia 8` generator.

Approach I

Event generator tuning using Bayesian optimization

Philip Ilten, Mike Williams, and Yunjie Yang

Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA 02139

- two histogram for each parameter: data_i and MC_i;
- Bayesian Optimization on the objective:

$$\chi^2 = \sum_{i=1}^{n_{bins}} \frac{(\mathrm{data}_i - \mathrm{MC}_i)^2}{\sigma^2_{\mathrm{data},i} + \sigma^2_{\mathrm{MC},i}}$$

- additional assumptions on distributions are required to guarantee convergence;

Approach II

Adversarial Variational Optimization of Non-Differentiable Simulators

Gilles Louppe G.LOUPPE@ULIEGE.BE
University of Liège, Belgium

Kyle Cranmer KYLE.CRANMER@NYU.EDU
New York University, USA

- an adversarial objective:

$$\text{Wasserstein}(F_{\text{real}}, F_\theta) = \sup_{d \in L_1} \mathop{\mathbb{E}}_{x \sim F_{\text{real}}} d(x) - \mathop{\mathbb{E}}_{x \sim F_\theta} d(x)$$

- Variational Optimization to search for distribution over **generator parameters**.

Assumptions and goals

We consider Adversarial Bayesian Optimization:

- no additional restrictions on distribution shapes;

Our primary concern is time complexity:

- sampling from the target event generator is expensive;
    - number of generator calls dominates overall complexity;
    - **minimizing number of event generator calls**;

- there is a configuration of generator that perfectly matches 'real' data.

Adversarial Bayesian Optimization

Adversarial Objective

Jensen-Shannon distance:

$$\mathrm{JS}(P, Q) = \log 2 + \frac{1}{2}\left[\mathop{\mathbb{E}}_{x\sim P} \log \frac{P(x)}{P(x) + Q(x)} + \mathop{\mathbb{E}}_{x\sim Q} \log \frac{Q(x)}{P(x) + Q(x)}\right] =$$

$$\log 2 - \min_f \text{cross-entropy}(f, P, Q)$$

- Jensen-Shannon distance can be approximated by a classifier.

Multi-Stage Adversarial Bayesian Optimization

- sequence of classifier models with increasing power:

$$\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_m = \mathcal{F}$$

- classifier $\mathcal{F}_i$ associated with 'pseudo' JS distance:

$$\mathrm{pJS}_i(P, Q) = \log 2 - \min_{f \in \mathcal{F}_i} \text{cross-entropy}(f, P, Q)$$

$$\mathrm{pJS}_1(P, Q) \leq \mathrm{pJS}_2(P, Q) \leq \cdots \leq \mathrm{pJS}_m(P, Q) = \mathrm{JS}(P, Q);$$

$$\boxed{\mathrm{pJS}_i(P, Q) \geq 0 \implies \mathrm{pJS}_{i+1}(P, Q) \geq 0}$$

Multi-Stage Adversarial Bayesian Optimization

$$\boxed{\mathrm{pJS}_i(P, Q) \geq 0 \implies \mathrm{pJS}_{i+1}(P, Q) \geq 0}$$

- ’weak’ classifiers tend to require less samples;
- ’weak’ classifiers can be used to rapidly explore search space;
- these results are constraints for a more powerful classifier.

Multi-Stage Adversarial Bayesian Optimization

$\mathrm{model}_1 = $ unconstrained BO on $\mathrm{pJS}_1(\mathrm{data}, \mathrm{generator}_\theta)$
for $k = 2, \ldots, m$ **do**
    $\mathrm{constraint}_k(\theta) = P\left(\mathrm{pJS}_{k-1} \leq 0 \;\mid\; \theta, \mathrm{model}_{k-1}\right)$
    $\mathrm{model}_k = $ BO on $\mathrm{pJS}_k(\mathrm{data}, \cdot)$ s.t. $\mathrm{constraint}_{\mathrm{j}}(\mathrm{theta}) > \tau$, $j = 0, \ldots, k-1$
end for

Experiments

Experiment

We follow problem statement from *Ilten P, Williams M, Yang Y. Event generator tuning using Bayesian optimization. Journal of Instrumentation. 2017 Apr 27;12(04):P04028.*

- e^+e^- modeled by `Pythia 8`;
- values of Monash tune as parameters of the 'real' distribution;

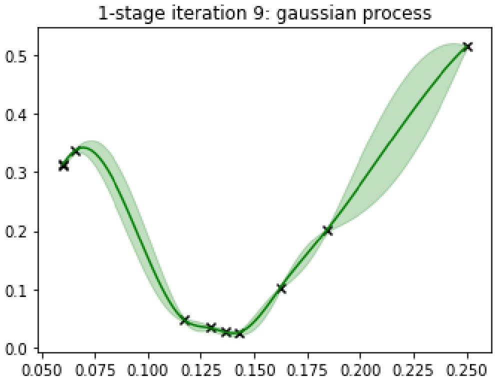
- 2-stage Adversarial Bayesian Optimization;
- number of samples required to avoid overfitting of the classifier is measured.

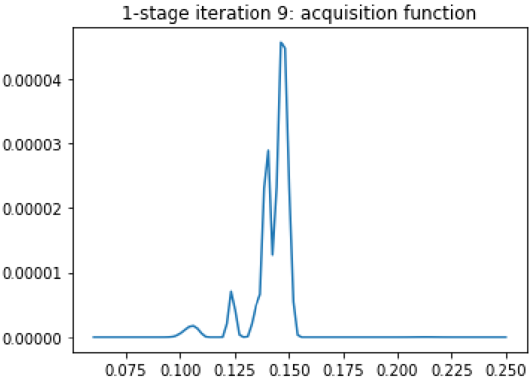
Experiment 1

Target generator options:

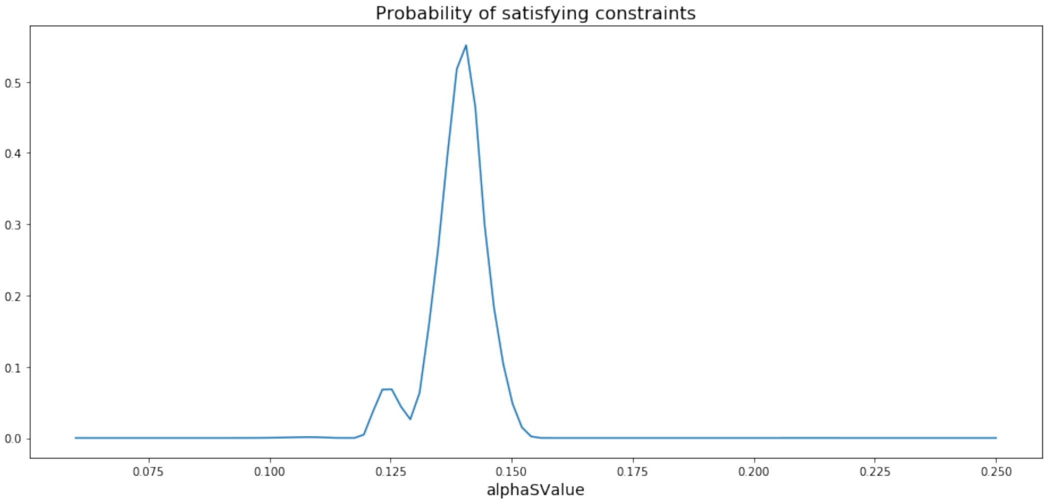
- `alphaSvalue`.

Experiment 1: stage 1

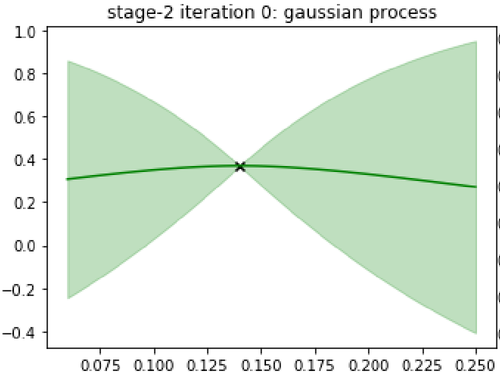


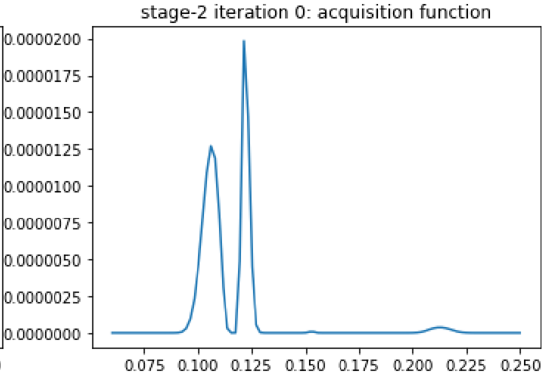


Experiment 1: stage 1

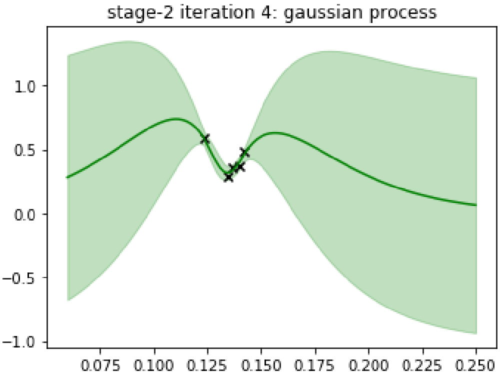


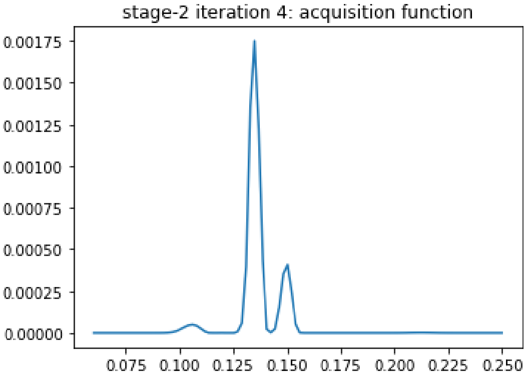
Experiment 1: stage 2



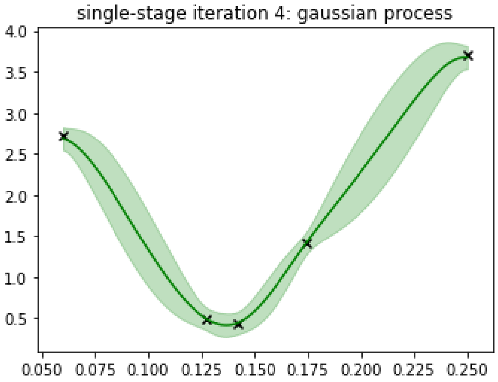


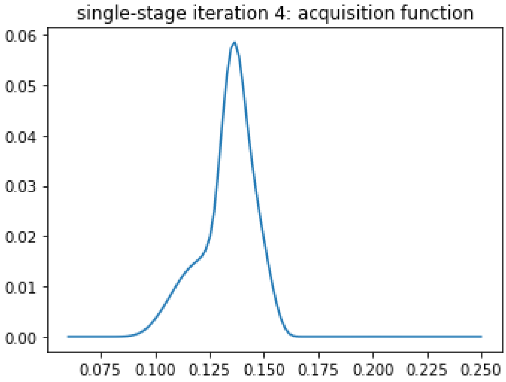
Experiment 1: stage 2



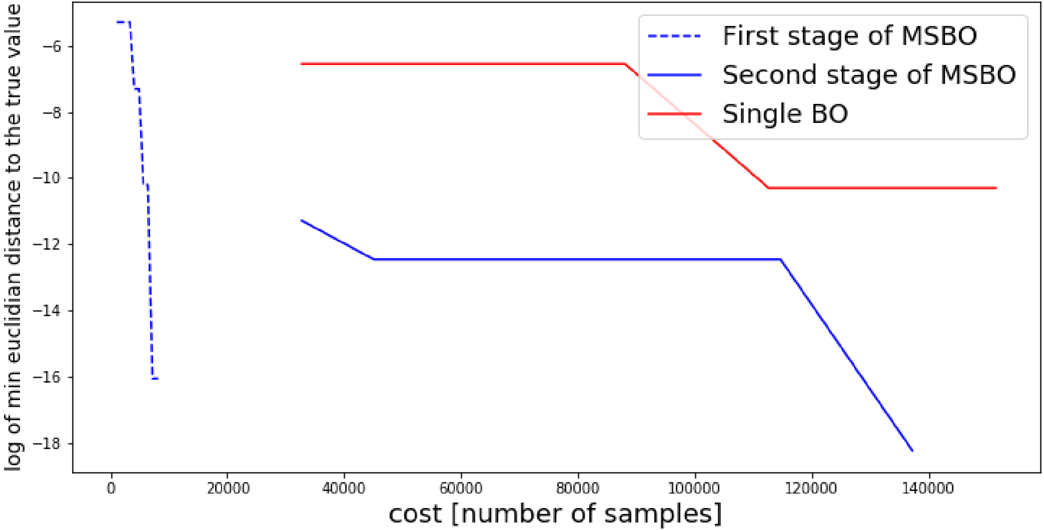


Experiment 1: single stage





Experiment 1: results



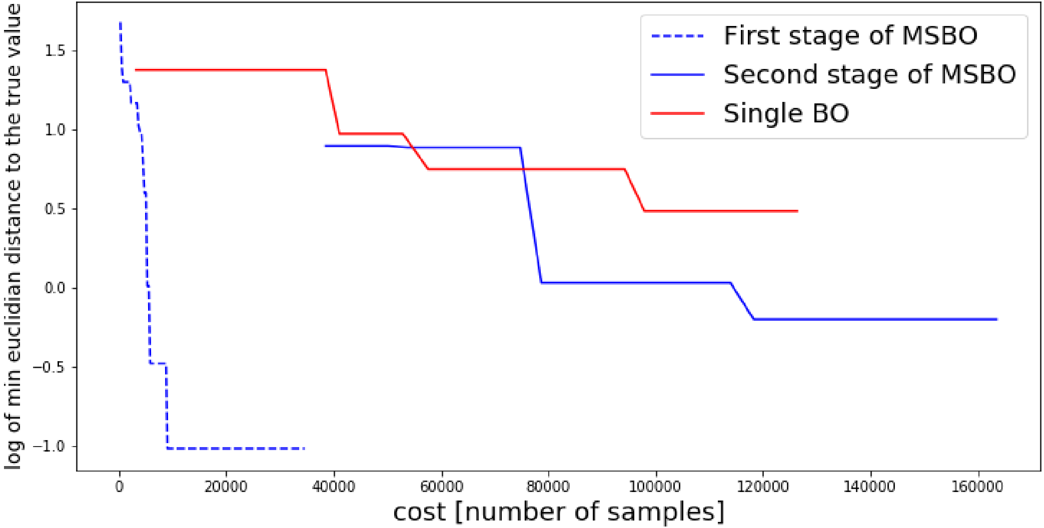
Experiment 2

Target generator options:

- `bLund`;
- `sigma`;
- `aExtraSQuark`;
- `aExtraDiQuark`;
- `rFactC`;
- `rFactB`.

Second group of varables from Ilten P, Williams M, Yang Y. Event generator tuning using Bayesian optimization. Journal of Instrumentation. 2017 Apr 27;12(04):P04028.

Experiment 2: results



Summary

Summary

- Adversarial Bayesian Optimization is a promising tool for tuning event generators;
- Multi-stage Adversarial Bayesian Optimization utilizes ’weak’ classifiers to incrementally constrain search space:
  - rapid exploration of search space on first stages;
  - late stages search for solution only among promising candidates;
  - **reduction in overall cost of optimization.**

Backup

Bayesian Adversarial Optimization

initialize Bayesian Optimization
while not bored **do**
    $\theta \leftarrow \mathrm{askBO}()$
    $X^{\theta}_{\mathrm{train}}, X^{\theta}_{\mathrm{test}} \leftarrow \mathrm{sample}(\theta)$
    $f \leftarrow$ train discriminator on $X^{\theta}_{\mathrm{train}}$ and $X^{\mathrm{real}}_{\mathrm{train}}$
    $\mathcal{L} \leftarrow \frac{1}{2 \cdot m} \left[\sum_{i=1}^{m} \log f(X^{\theta,i}_{\mathrm{test}}) + \sum_{i=1}^{m} \log(1 - f(X^{\mathrm{real},i}_{\mathrm{test}})) \right]$
    $\mathrm{tellBO}(\theta, \log 2 - \mathcal{L})$
end while

Possible Caveats

- constraints are observed by authors to mess with GP;
- without assumption $\exists \theta \ : \ \mathrm{JS}(\mathrm{generator}(\theta), \mathrm{real}) = 0$:
    - it is likely that the method would still work (modifying constraints) if classifiers are from the same family of algorithms;
    - it is possible, that BO with weak classifier carries no information about BO with a strong classifier.

Expected Improvement with Constraints

Problem:

$$\begin{aligned} \mathrm{EI}(x) \quad &\rightarrow \quad \min; \\ \text{s.t.} \quad &g(x) \geq 0. \end{aligned}$$

- improvement is impossible if constraints are violated:

$$\mathrm{CEI}(x) = P(g(x) \geq 0) \cdot \mathrm{EI}(x) + P(g(x) < 0) \cdot 0$$

- constraints in our case: $\mathrm{model}_i(x) \leq 0$.

Gelbart, M.A., Snoek, J. and Adams, R.P., 2014. Bayesian optimization with unknown constraints. arXiv preprint arXiv:1403.5607.

Technical details

- training set is incrementally extended until over-fitting becomes insignificant.
- 2 stage ABO:
  - 1 stage: XGboost with 1 tree and max depth = 3;
  - 2 stage: XGboost with 20 tree and max depth = 6.

Experiment 1